Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/004513

International filing date: 11 February 2005 (11.02.2005)

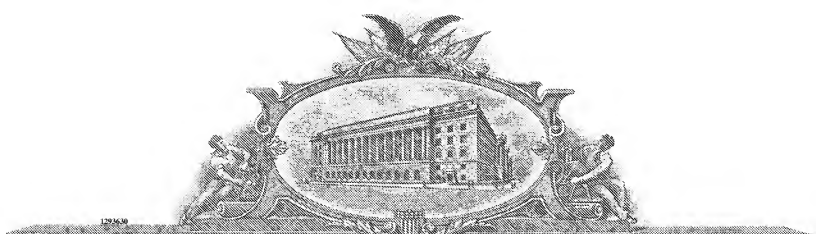
Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/544,788
Filing date: 13 February 2004 (13.02.2004)

Date of receipt at the International Bureau: 17 March 2005 (17.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse



129630

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

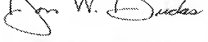
United States Patent and Trademark Office

March 08, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/544,788***
FILING DATE: ***February 13, 2004***
RELATED PCT APPLICATION NUMBER: ***PCT/US05/04513***

Certified by



Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

To type a plus sign (+) in this box [+]

Docket Number **DNA1180**

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

INVENTOR(s)/APPLICANT(s)

Given Name (first and middle [if any])	Family Name or Surname	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Tony N.	Frudakis	Bradenton, Florida

☐ Additional inventors are being named on page ____ attached hereto.

TITLE OF INVENTION (280 characters max)

METHODS AND COMPOSITIONS FOR INFERRING EYE COLOR

CORRESPONDENCE ADDRESS

LISA A. HAILE
GRAY CARY WARE & FREIDENRICH
4365 Executive Drive, Suite 1100, San Diego, California 92121-2133
Phone: (858) 677-1456 Fax: (858) 677-1465

ENCLOSED APPLICATION PARTS (*check all that apply*)

X	Specification	*Number of Pages*	21		X	Entitled to Small Entity Status	Yes
X	Drawing(s)	*Number of Sheets*	2		X	Other (*specify*)	Specification Cover Page (1 pg.) Return Receipt Postcard

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (*check one*)

X	A check or money order is enclosed to cover the filing fees		FILING FEE AMOUNT
X	The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: A duplicate copy of this sheet is enclosed.	**50-1355**	$80.00 Check # 53729

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No.

Yes, the name of the U.S. Government agency and the Government contract number are:

Respectfully submitted

SIGNATURE _____ DATE February 13, 2004

Richard J. Imbra, Reg. No. 37,643
TYPED or PRINTED NAME, Registration No.

TELEPHONE (858) 677-1496

Certificate of Mailing by Express Mail

Express Mail Label No. **EV 318 741 156 US** Date February 13, 2004

I certify that this provisional patent application cover sheet, provisional patent application and fee are being deposited with sufficient postage on February 13, 2004 with the U.S. Postal Service as "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above in an envelope addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

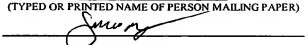
Saul Solano

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"EXPRESS MAIL LABEL NUMBER: EV 318 741 156 US

I certify that the attached provisional patent application is being deposited with sufficient postage on February 13, 2004 with the U.S. Postal Service as "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above in an envelope addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

SAUL SOLANO
(TYPED OR PRINTED NAME OF PERSON MAILING PAPER)



(SIGNATURE OF PERSON MAILING PAPER AND FEE)

PROVISIONAL
APPLICATION

for

UNITED STATES LETTERS PATENT

on

METHODS AND COMPOSITIONS FOR INFERRING EYE COLOR

by

Tony N. Frudakis

Docket No.: DNA1180
Specification: 21 pages
Drawings: 2 Sheets

Lisa A. Haile, Ph.D.
Gray Cary Ware & Freidenrich LLP
4365 Executive Drive, Suite 1100
San Diego, California 92121-2133

METHODS AND COMPOSITIONS FOR INFERRING EYE COLOR

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0001] The invention relates generally to methods of determining eye color of an individual, and more specifically to methods of inferring eye color of an individual by identifying single nucleotide polymorphisms (SNPs) associated with eye color in a nucleic acid sample of the individual, and to compositions useful for practicing such methods.

BACKGROUND INFORMATION

[0002] Biotechnology has revolutionized the field of forensics. More specifically, the identification of polymorphic regions in human genomic DNA has provided a means to distinguish individuals based on the occurrence of a particular nucleotide at each of several positions in the genomic DNA that are known to contain polymorphisms. As such, analysis of DNA from an individual allows a genetic fingerprint or "bar code" to be constructed that, with the possible exception of identical twins, essentially is unique to one particular individual in the entire human population.

[0003] In combination with DNA amplification methods, which allow a large amount of DNA to be prepared from a sample as small as a spot of blood or semen or a hair follicle, DNA analysis has become a routine tool in criminal cases as evidence that can free or, in some cases, convict a suspect. Indeed, criminal courts, which do not yet allow the results of a lie detector test into evidence, admit DNA evidence into trial. In addition, DNA extracted from evidence that, in some cases, has been preserved for years after the crime was committed, has resulted in the convictions of many people being overturned.

[0004] Although DNA fingerprint analysis has greatly advanced the field of forensics, and has resulted in freedom of people, who, in some cases, were erroneously imprisoned for years, current DNA analysis methods are limited. In particular, DNA fingerprinting analysis only provides confirmatory evidence that a particular person is, or is not, the person from which the sample was derived. For example, while DNA in a semen sample can be used to obtain a specific "bar code", it provides no information about the person that left the sample. Instead, the bar code can only be compared to the bar code of a suspect in the

crime. If the bar codes match, then it can reasonably be concluded that the person likely is the source of the semen. However, if there is not a match, the investigation must continue.

[0005] An effort has begun to accumulate a database of bar codes, particularly of convicted criminals. Such a database allows prospective use of a bar code obtained from a biological sample left at a crime scene; i.e., the bar code of the sample can be compared, using computerized methods, to the bar codes in the database and, where the sample is that of a person whose bar code is in the database, a match can be obtained, thus identifying the person as the likely source of the sample from the crime scene. While the availability of such a database provides a significant advance in forensic analysis, the potential of DNA analysis is still limited by the requirement that the database must include information relating to the person who left the biological sample at the crime scene, and it likely will be a long time, if ever, that such a database will provide information of an entire population. Thus, there is a need for methods that can provide prospective information about a subject from a nucleic acid sample of the subject.

SUMMARY OF THE INVENTION

[0006] The present invention provides methods of inferring the eye color of a human subject from a nucleic acid sample or a polypeptide sample of the subject, and compositions for practicing such methods. The methods of the invention are based, in part, on the identification of single nucleotide polymorphisms (SNPs) that, alone or in combination, allow an inference to be drawn as to eye shade or eye color. As such, the methods can utilize the identification of haploid or diploid alleles of SNPs and or haplotypes. The compositions and methods of the invention are useful, for example, as forensic tools for obtaining information relating to physical characteristics of a potential crime victim or a perpetrator of a crime from a nucleic acid sample present at a crime scene, and as tools to assist in breeding domesticated animals, livestock, and the like to contain a pigmentation trait as desired.

[0007] In one embodiment, the invention relates to a method of inferring eye color of a human individual by determining the nucleotide occurrence of at least one SNP as set forth in Table 1 (see, also, Table 2; SEQ ID NOS:1 to 35). In particular, the method comprises

determining the nucleotide occurrence of at least one SNP as set forth in any of SEQ ID NOS:1 to 3, 7 to 9, 11 to 13, 15 to 18, 20, 22 to 31, and 35. In one aspect of this embodiment, the method comprises identifying at least two nucleotide occurrences of the SNP position, including, for example, diploid alleles corresponding to at least one SNP position, or a haplotype corresponding to at least two SNP positions.

[0008] In another embodiment, the present invention relates to compositions useful for sampling a nucleic acid sample to determine a nucleotide occurrence of at least one SNP. Such compositions include, for example, oligonucleotide probes that selectively hybridize to a nucleic acid molecule including one or the other of a nucleotide occurrence of a SNP (e.g., a nucleic acid molecule containing either a "G" or an "A" residue at the SNP position of SEQ ID NO:1 (see, also, Table 3; marker 2142); or oligonucleotide primers that selectively hybridize to a position upstream or downstream (or both) of the nucleotide position such that a primer extension reaction or a nucleic acid amplification reaction can generate a product including the SNP position. Where the nucleotide occurrence of a SNP position is in a gene coding sequence, and the alternative forms of the SNP result in a change in the encoded amino acid, the composition for detecting the nucleotide occurrence at the SNP position can be an antibody that specifically binds to a polypeptide containing one or the other amino acid residue, but not to both such polypeptides.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Figure 1 shows the distribution of eye color scores determined as described in Example 1.

[0010] Figure 2 shows the distribution of eye color related SNPs along the human chromosomes. Dots indicate known human pigmentation genes, and dashes represent the most strongly associated of the selected SNPs (27 shown; see Example 1).

DETAILED DESCRIPTION OF THE INVENTION

[0011] The present invention is based, in part, an the identification of a panel of single nucleotide polymorphisms (SNPs) that alone, or in combinations, allow an inference to be drawn as to the eye color of an individual from a nucleic acid or protein sample of the

individual. As disclosed herein, many of these SNPs came from a pan-genome screen and are dispersed among the chromosomes (see Figure 2). As such the SNPs can be used individually, and in combinations, including as haploid or diploid alleles, to draw an inference regarding eye color. In addition, where the SNPs are present in the same gene or are sufficiently linked, they can be assembled into haplotypes, and haploid and/or diploid haplotype alleles can be used to infer eye color.

[0012] The term "haplotype" is used herein to refer to groupings of two or more nucleotide SNPs that are linked. As such, the SNPs can be present in the same gene or in adjacent genes or in a gene and an adjacent intergenic region, or otherwise present in the genome such that they segregate non-randomly. The term "haplotype alleles" as used herein refers to a non-random combination of nucleotide occurrences of SNPs that make up a haplotype.

[0013] The term "penetrant pigmentation-related haplotype alleles" refers to haplotype alleles whose association with eye color pigmentation is strong enough that it can be detected using simple genetics approaches. Corresponding haplotypes of penetrant pigmentation-related haplotype alleles, are referred to herein as "penetrant pigmentation-related haplotypes." Similarly, individual nucleotide occurrences of SNPs are referred to herein as "penetrant pigmentation-related SNP nucleotide occurrences" if the association of the nucleotide occurrence with the eye color pigmentation trait is strong enough on its own to be detected using simple genetics approaches, or if the SNP loci for the nucleotide occurrence make up part of a penetrant haplotype. The corresponding SNP loci are referred to herein as "penetrant pigmentation-related SNPs." Haplotype alleles of penetrant haplotypes are also referred to herein as "penetrant haplotype alleles" or "penetrant genetic features." Penetrant haplotypes are also referred to herein as "penetrant genetic feature SNP combinations.

[0014] The term "latent pigmentation-related haplotype alleles" refers to haplotype alleles that, in the context of one or more penetrant haplotypes, strengthen the inference of the genetic eye color pigmentation trait. Latent pigmentation-related haplotype alleles are typically alleles whose association with eye color pigmentation is not strong enough to be

detected with simple genetics approaches. Latent pigmentation-related SNPs are individual SNPs that make up latent pigmentation-related haplotypes.

[0015] A sample useful for practicing a method of the invention can be any biological sample of a subject that contains nucleic acid molecules, including portions of the gene sequences to be examined, or corresponding encoded polypeptides, depending on the particular method. As such, the sample can be a cell, tissue or organ sample, or can be a sample of a biological fluid such as semen, saliva, blood, and the like. A nucleic acid sample useful for practicing a method of the invention will depend, in part, on whether the SNPs to be identified are in coding regions or in non-coding regions. Thus, where at least one of the SNPs to be identified is in a non-coding region, the nucleic acid sample generally is a deoxyribonucleic acid (DNA) sample, particularly genomic DNA or an amplification product thereof. However, where heteronuclear ribonucleic acid (RNA), which includes unspliced mRNA precursor RNA molecules, is available, a cDNA or amplification product thereof can be used. Where the each of the SNPs is present in a coding region of the pigmentation gene(s), the nucleic acid sample can be DNA or RNA, or products derived therefrom, for example, amplification products. Furthermore, while the methods of the invention generally are exemplified with respect to a nucleic acid sample, it will be recognized that particular SNP alleles can be in coding regions of a gene and can result in polypeptides containing different amino acids at the positions corresponding to the SNPs due to non-degenerate codon changes. As such, in one aspect, the methods of the invention can be practiced using a sample containing polypeptides of the subject.

[0016] Methods of the invention can be practiced with respect to human subjects and, therefore, can be particularly useful for forensic analysis. In a forensic application or a method of the invention, the human nucleic acid sample can be obtained from a crime scene, using well established sampling methods. Thus, the sample can be fluid sample or a swab sample. For example, the sample can be a swab sample, blood stain, semen stain, hair follicle, or other biological specimen, taken from a crime scene, or can be a soil sample suspected of containing biological material of a potential crime victim or perpetrator, can be material retrieved from under the finger nails of a potential crime victim, or the like,

wherein nucleic acids (or polypeptides) in the sample can be used as a basis for drawing an inference as to eye color according to a method of the invention.

[0017] A mammalian subject that can be examined according to a method of the invention can be any mammalian species. In particular, the methods are applicable to drawing an inference as to a pigmentation trait of a human subject. The human subject can be from a general population of mixed ethnicity, or the human subject can be of a particular ethnic background or race. For example, the subject can be a Caucasian. With respect to non-human mammalian species, the methods of the invention are valuable in providing predictions of commercially valuable eye color phenotypes, for example, in breeding.

[0018] The sequences disclosed in Table 3 provide flanking nucleotide sequences for the SNPs disclosed herein. These flanking sequence serve to aid in the identification of the precise location of the SNPs in the human genome, and serve as target gene segments useful for performing methods of the invention. A target polynucleotide typically includes a SNP locus and a segment of a corresponding gene that flanks the SNP. Primers and probes that selectively hybridize at or near the target polynucleotide sequence, as well as specific binding pair members that can specifically bind at or near the target polynucleotide sequence, can be designed based on the disclosed gene sequences and information provided herein.

[0019] As used herein, the term "selective hybridization" or "selectively hybridize," refers to hybridization under moderately stringent or highly stringent conditions such that a nucleotide sequence preferentially associates with a selected nucleotide sequence over unrelated nucleotide sequences to a large enough extent to be useful in identifying a nucleotide occurrence of a SNP. It will be recognized that some amount of non-specific hybridization is unavoidable, but is acceptable provided that hybridization to a target nucleotide sequence is sufficiently selective such that it can be distinguished over the non-specific cross-hybridization, for example, at least about 2-fold more selective, generally at least about 3-fold more selective, usually at least about 5-fold more selective, and particularly at least about 10-fold more selective, as determined, for example, by an amount of labeled oligonucleotide that binds to target nucleic acid molecule as compared to a

nucleic acid molecule other than the target molecule, particularly a substantially similar (i.e., homologous) nucleic acid molecule other than the target nucleic acid molecule. Conditions that allow for selective hybridization can be determined empirically, or can be estimated based, for example, on the relative GC:AT content of the hybridizing oligonucleotide and the sequence to which it is to hybridize, the length of the hybridizing oligonucleotide, and the number, if any, of mismatches between the oligonucleotide and sequence to which it is to hybridize (see, for example, Sambrook et al., "Molecular Cloning: A laboratory manual (Cold Spring Harbor Laboratory Press 1989)).

[0020] An example of progressively higher stringency conditions is as follows: 2 x SSC/0.1% SDS at about room temperature (hybridization conditions); 0.2 x SSC/0.1% SDS at about room temperature (low stringency conditions); 0.2 x SSC/0.1% SDS at about 42°C (moderate stringency conditions); and 0.1 x SSC at about 68°C (high stringency conditions). Washing can be carried out using only one of these conditions, e.g., high stringency conditions, or each of the conditions can be used, e.g., for 10-15 minutes each, in the order listed above, repeating any or all of the steps listed. However, as mentioned above, optimal conditions will vary, depending on the particular hybridization reaction involved, and can be determined empirically.

[0021] The term "polynucleotide" is used broadly herein to mean a sequence of deoxyribonucleotides or ribonucleotides that are linked together by a phosphodiester bond. For convenience, the term "oligonucleotide" is used herein to refer to a polynucleotide that is used as a primer or a probe. Generally, an oligonucleotide useful as a probe or primer that selectively hybridizes to a selected nucleotide sequence is at least about 15 nucleotides in length, usually at least about 18 nucleotides, and particularly about 21 nucleotides or more in length.

[0022] A polynucleotide can be RNA or can be DNA, which can be a gene or a portion thereof, a cDNA, a synthetic polydeoxyribonucleic acid sequence, or the like, and can be single stranded or double stranded, as well as a DNA/RNA hybrid. In various embodiments, a polynucleotide, including an oligonucleotide (e.g., a probe or a primer) can contain nucleoside or nucleotide analogs, or a backbone bond other than a phosphodiester

bond. In general, the nucleotides comprising a polynucleotide are naturally occurring deoxyribonucleotides, such as adenine, cytosine, guanine or thymine linked to 2'-deoxyribose, or ribonucleotides such as adenine, cytosine, guanine or uracil linked to ribose. However, a polynucleotide or oligonucleotide also can contain nucleotide analogs, including non-naturally occurring synthetic nucleotides or modified naturally occurring nucleotides. Such nucleotide analogs are well known in the art and commercially available, as are polynucleotides containing such nucleotide analogs (Lin et al., *Nucl. Acids Res.* 22:5220-5234 (1994); Jellinek et al., *Biochemistry* 34:11363-11372 (1995); Pagratis et al., *Nature Biotechnol.* 15:68-73 (1997), each of which is incorporated herein by reference).

[0023] The covalent bond linking the nucleotides of a polynucleotide generally is a phosphodiester bond. However, the covalent bond also can be any of numerous other bonds, including a thiodiester bond, a phosphorothioate bond, a peptide-like bond or any other bond known to those in the art as useful for linking nucleotides to produce synthetic polynucleotides (see, for example, Tam et al., *Nucl. Acids Res.* 22:977-986 (1994); Ecker and Crooke, *BioTechnology* 13:351360 (1995), each of which is incorporated herein by reference). The incorporation of non-naturally occurring nucleotide analogs or bonds linking the nucleotides or analogs can be particularly useful where the polynucleotide is to be exposed to an environment that can contain a nucleolytic activity, including, for example, a tissue culture medium or upon administration to a living subject, since the modified polynucleotides can be less susceptible to degradation.

[0024] A polynucleotide or oligonucleotide comprising naturally occurring nucleotides and phosphodiester bonds can be chemically synthesized or can be produced using recombinant DNA methods, using an appropriate polynucleotide as a template. In comparison, a polynucleotide or oligonucleotide comprising nucleotide analogs or covalent bonds other than phosphodiester bonds generally are chemically synthesized, although an enzyme such as T7 polymerase can incorporate certain types of nucleotide analogs into a polynucleotide and, therefore, can be used to produce such a polynucleotide recombinantly from an appropriate template (Jellinek et al., *supra*, 1995). Thus, the term polynucleotide as used herein includes naturally occurring nucleic acid molecules, which can be isolated from

a cell, as well as synthetic molecules, which can be prepared, for example, by methods of chemical synthesis or by enzymatic methods such as by the polymerase chain reaction (PCR).

[0025] In various embodiments, it can be useful to detectably label a polynucleotide or oligonucleotide. Detectable labeling of a polynucleotide or oligonucleotide is well known in the art. Particular non-limiting examples of detectable labels include chemiluminescent labels, radiolabels, enzymes, haptens, or even unique oligonucleotide sequences.

[0026] A method of the identifying an eye color related SNP also can be performed using a specific binding pair member. As used herein, the term "specific binding pair member" refers to a molecule that specifically binds or selectively hybridizes to another member of a specific binding pair. Specific binding pair member include, for example, probes, primers, polynucleotides, antibodies, etc. For example, a specific binding pair member can be a primer or a probe that selectively hybridizes to a target polynucleotide that includes a SNP locus, or that hybridizes to an amplification product generated using the target polynucleotide as a template, or can be an antibody that, under the appropriate conditions, selectively binds to a polypeptide containing one, but not the other, variant encoded by a polynucleotide comprising a particular SNP.

[0027] Numerous methods are known in the art for determining the nucleotide occurrence for a particular SNP in a sample. Such methods can utilize one or more oligonucleotide probes or primers, including, for example, an amplification primer pair, that selectively hybridize to a target polynucleotide, which contains one or more pigmentation-related SNP positions. Oligonucleotide probes useful in practicing a method of the invention can include, for example, an oligonucleotide that is complementary to and spans a portion of the target polynucleotide, including the position of the SNP, wherein the presence of a specific nucleotide at the position (i.e., the SNP) is detected by the presence or absence of selective hybridization of the probe. Such a method can further include contacting the target polynucleotide and hybridized oligonucleotide with an endonuclease, and detecting the presence or absence of a cleavage product of the probe, depending on whether the

nucleotide occurrence at the SNP site is complementary to the corresponding nucleotide of the probe.

[0028] An oligonucleotide ligation assay also can be used to identify a nucleotide occurrence at a polymorphic position, wherein a pair of probes that selectively hybridize upstream and adjacent to and downstream and adjacent to the site of the SNP, and wherein one of the probes includes a terminal nucleotide complementary to a nucleotide occurrence of the SNP. Where the terminal nucleotide of the probe is complementary to the nucleotide occurrence, selective hybridization includes the terminal nucleotide such that, in the presence of a ligase, the upstream and downstream oligonucleotides are ligated. As such, the presence or absence of a ligation product is indicative of the nucleotide occurrence at the SNP site.

[0029] An oligonucleotide also can be useful as a primer, for example, for a primer extension reaction, wherein the product (or absence of a product) of the extension reaction is indicative of the nucleotide occurrence. In addition, a primer pair useful for amplifying a portion of the target polynucleotide including the SNP site can be useful, wherein the amplification product is examined to determine the nucleotide occurrence at the SNP site. Particularly useful methods include those that are readily adaptable to a high throughput format, to a multiplex format, or to both. The primer extension or amplification product can be detected directly or indirectly and/or can be sequenced using various methods known in the art. Amplification products which span a SNP loci can be sequenced using traditional sequence methodologies (e.g., the "dideoxy-mediated chain termination method," also known as the "Sanger Method"(Sanger, F., et al., *J. Molec. Biol.* 94:441 (1975); Prober et al. *Science* 238:336-340 (1987)) and the "chemical degradation method," "also known as the "Maxam-Gilbert method"(Maxam, A. M., et al., *Proc. Natl. Acad. Sci.* (U.S.A.) 74:560 (1977)), both references herein incorporated by reference) to determine the nucleotide occurrence at the SNP loci.

[0030] Methods of the invention can identify nucleotide occurrences at SNPs using a "microsequencing" method. Microsequencing methods determine the identity of only a single nucleotide at a "predetermined" site. Such methods have particular utility in

determining the presence and identity of polymorphisms in a target polynucleotide. Such microsequencing methods, as well as other methods for determining the nucleotide occurrence at a SNP loci are discussed in Boyce-Jacino et al., U.S. Pat. No. 6,294,336, which is incorporated herein by reference.

[0031] Microsequencing methods include the Genetic Bit Analysis method disclosed by Goelet, P. et al. (WO 92/15712, herein incorporated by reference). Additional, primer-guided, nucleotide incorporation procedures for assaying polymorphic sites in DNA have also been described (Komher et al, Nucl. Acids. Res. 17:7779-7784 (1989); Sokolov, Nucl. Acids Res. 18:3671 (1990); Syvanen et al., Genomics 8:684-692 (1990); Kuppuswamy et al., Proc. Natl. Acad. Sci. (U.S.A.) 88:1143-1147 (1991); Prezant et al, Hum. Mutat. 1:159-164 (1992); Ugozzoli et al., GATA 9:107-112 (1992); Nyren et al., Anal. Biochem. 208:171-175 (1993); and Wallace, WO89/10414). These methods differ from Genetic Bit™ analysis in that they all rely on the incorporation of labeled deoriboxynucleotides to discriminate between bases at a polymorphic site. In such a format, since the signal is proportional to the number of deoriboxynucleotides incorporated, polymorphisms that occur in runs of the same nucleotide can result in signals that are proportional to the length of the run (Syvanen et al. Amer. J. Hum. Genet. 52:46-59 (1993)). Alternative microsequencing methods have been provided by Mundy, (U.S. Pat. No. 4,656,127) and Cohen et al (French Patent 2,650,840; PCT Appl. No. WO91/02087) which discusses a solution-based method for determining the identity of the nucleotide of a polymorphic site. As in the Mundy method of U.S. Pat. No. 4,656,127, a primer is employed that is complementary to allelic sequences immediately 3'-to a polymorphic site.

[0032] In response to the difficulties encountered in employing gel electrophoresis to analyze sequences, alternative methods for microsequencing have been developed. Macevicz (U.S. Pat. No. 5,002,867), for example, describes a method for determining nucleic acid sequence via hybridization with multiple mixtures of oligonucleotide probes. In accordance with such method, the sequence of a target polynucleotide is determined by permitting the target to sequentially hybridize with sets of probes having an invariant nucleotide at one position, and a variant nucleotides at other positions. The Macevicz

method determines the nucleotide sequence of the target by hybridizing the target with a set of probes, and then determining the number of sites that at least one member of the set is capable of hybridizing to the target (i.e., the number of "matches"). This procedure is repeated until each member of a sets of probes has been tested. Boyce-Jacino et al. (U.S. Pat. No. 6,294,336) provide a solid phase sequencing method for determining the sequence of nucleic acid molecules (either DNA or RNA) by utilizing a primer that selectively binds a polynucleotide target at a site wherein the SNP is the most 3' nucleotide selectively bound to the target.

[0033] In one particular commercial example of a method that can be used to identify a nucleotide occurrence of one or more SNPs, the nucleotide occurrences of pigmentation-related SNPs in a sample can be determined using the SNP-IT™ method (Orchid BioSciences, Inc.; Princeton, NJ). In general, the SNP-IT™ method is a 3-step primer extension reaction. In the first step a target polynucleotide is isolated from a sample by hybridization to a capture primer, which provides a first level of specificity. In a second step the capture primer is extended from a terminating nucleotide trisphosphate at the target SNP site, which provides a second level of specificity. In a third step, the extended nucleotide trisphosphate can be detected using a variety of known formats, including: direct fluorescence, indirect fluorescence, an indirect colorimetric assay, mass spectrometry, fluorescence polarization, etc. Reactions can be processed in 384 well format in an automated format using a SNPstream™ instrument (Orchid BioSciences, Inc.). Phase known data can be generated by inputting phase unknown raw data from the SNPstream™ instrument into the Stephens and Donnelly's PHASE program.

[0034] The method of identifying a nucleotide occurrence in the sample for at least one eye color related SNP, as discussed above, can further include grouping the nucleotide occurrences of the SNPs into one or more haplotype alleles indicative of eye color. To infer eye color of a test subject, the identified haplotype alleles can be compared to known haplotype alleles, wherein the relationship of the known haplotype alleles to eye color is known.

[0035] The following example is intended to illustrate but not limit the invention.

EXAMPLE 1

IDENTIFICATION OF SNPs INDICATIVE OF EYE COLOR

[0036] This example describes the identification of SNPs useful for inferring eye color from a nucleic acid sample of an individual.

[0037] Iris colors were measured using a Cannon digital camera. Each subject peered into a cardboard box at one end, and the camera at the other end took the photo under a standardized brightness from a constant distance for each; 100 samples were collected using this method. Adobe Photoshop™ software was used to quantify the luminosity and the red/green, green/blue and red/blue wavelength reflectance ratios for the left iris; lighter eye colors had lower values for each of these variables. For each variable, the scores were scaled about the mean value. For example an eye of the average red/green value received a new scaled value of 1, with those of value below the mean converted to values less than 1 (proportional to their difference from the mean) and those greater than the mean converted to values greater than 1 (proportional to their difference from the mean). The scaled red/green, red/blue and green/blue values were summed for each eye and added together. This value was added to a scaled luminosity value for each eye to produce an eye color score for that eye. The eye color scores showed a continuous distribution (see FIG. 1).

[0038] The lightest 21 (at the top of the above distribution) were selected, and pooled into a "Light" sample; and the darkest 21 eye color samples (at the bottom of the above distribution) were selected and pooled into a "Dark" sample. A GeneChip® Mapping 10K Array and Assay Set (Affymetrix; Santa Clara CA) was used to screen each pool. For each of the 10,000 SNPs on the GeneChip® array, an allele frequency was calculated for the Light pool and the Dark pool. The 10,000 SNPs were ranked based on the allele frequency differential between the two groups (Delta value), a Pearson's P value statistic, and an Odds Ratio statistic on the allele frequency differential between the two groups. The top 100 SNPs based on the Odds Ratio statistic were selected, as were all others that were in the top 100 for Delta value and Pearson's P value (even if not in the top 100 based on the Odds ratio test) to produce a set of 130 SNPs.

[0039] To validate which of the 130 SNPs were associated with iris colors, a second completely separate group of 100 samples was genotyped and ranked in the same way. The best 60 SNPs described in PCT/US02/16789, which is incorporated herein by reference, also were genotyped in this same sample of 100 subjects. Of the 190 candidate SNPs, approximately 30 showed either a good Delta value, Pearson's P value or Odds ratio test statistic. The distribution of the 30 selected SNPs along the chromosomes is shown in FIG. 2. Table 1 shows the delta value, chromosomal position for 27 of the SNPs, and indicates whether the SNP is located within a known pigmentation gene or within a few megabases (Mb) of a known pigmentation gene.

[0040] Those SNPs indicated as located "in OCA2" or "in "ASIP" or "in TYRP1" in the above list previously were identified, and are disclosed in PCT/US02/16789; their inclusion in the list of Table 1 provides confirmation of their value as disclosed in PCT/US02/16789. The remaining SNPs are newly disclosed herein, and were identified using the Affymetrix chip.

[0041] A classification model was built using the 27 SNPs listed in Table 1, whereby the 200 subjects used to discover them were classified into Light or Dark eye color groups. Neural nets gave a classification accuracy of about 95% within-model, and about 80% outside model. It is noted that neural nets generally require a much larger sample size for the number of variables used here. A simpler method was used to obtain a within-model accuracy of 97%.

[0042] Table 2 provides a list of 35 SNPs, including 15 of the 27 SNPs shown in Table 1, and 20 additional SNPs. The designation "unknown" or "V2-unknown" is used to identify SNPs that were not disclosed in PCT/US02/16789 (SEQ ID NOS:1 to 3, 7 to 9, 11 to 13, 16 to31 and 35; see, also, Table 3). The 20 additional SNPs in Table 2 were selected because they had interesting distributions that were helpful for classification analysis, but had less optimal P-values or delta values (Note: Table 1 has a cut-off Delta value of 0.125, whereas Table 2 includes 15 SNPs that also are in Table 1 (and have a Delta value greater than 0.125) as well as 20 SNPs having Delta values less than 0.125, but otherwise having an interesting distribution). For example, one of the SNPs in Table 2 had an interesting

distribution in that only 5 CT genotypes (the rest were CC genotypes; i.e. T is rare), but the T occurred in Light eyes every time. Thus, while its Delta value and P-value were not very good, the SNP was selected as having potential interest (stress potential).

[0043] Table 3 provides sequences that flank and include the SNPs listed in Table 2. Correspondence can be determined with reference to the "MARKER" number. The position of the SNP in the sequences is indicated in bold, and the alternative nucleotide occurrence are shown as ALL1 (Allele 1) and ALL2 (Allele 2). The gene and SNP names also are included. Additional flanking sequences can be determined by using the disclosed sequences to search a database such as GenBank (see, e.g., the National Center for Biotechnology Information, on the world wide web, URL "www.ncbi.nlm.nih.gov"). Based on these sequences, probes and primers, including primer pairs, can be designed for determining the nucleotide occurrence at a SNP position.

[0044] Although the invention has been described with reference to the above example, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention. Accordingly, the invention is limited only by the claims, which follow Tables 1 to 3.

TABLE 1

Marker	DELTA	Position	Pigment Gene
2142	0.275	Xp11.23	
2190	0.247619	12q12	
2121	0.215476	1q21-23	
2189	0.211905	Xp11.23	
1879	0.199248	15q11.2-12	in OCA2 (15q11.2)
1916	0.188095	15q11.2-12	in OCA2 (15q11.2)
1908	0.183333	15q11.2-12	in OCA2 (15q11.2)
2109	0.164286	1q25-31	
2177	0.157895	1q44	
2130	0.154762	13q12.3	
2191	0.15	3q23-q24	5Mb from HPS3 (3q23-q24)
2126	0.141667	6q22	
1998	0.136905	10q24	3Mb from HPS1 (10q23)
2110	0.136905	14q24.3	
2147	0.136905	12p11.2	
1876	0.132143	9p23	in TYRP1 (9p23)
2113	0.130952	4q28-31.1	
2201	0.129762	5p15.2	
1979	0.128571	20q11.2-q12	in ASIP (20q11.2-q12)
1986	0.128571	20q11.2-q12	in ASIP (20q11.2-q12)
2178	0.128571	1p13	
2050	0.126566	3q24	in HPS3 (3q23-q24)
2169	0.126316	13q31.5	2Mb from DCT (13q32)
1873	0.12619	15q11.2-q12	in OCA2 (15q11.2)
2168	0.12619	1p34.3	
2156	0.125	11p11.2	
2205	0.125	16p13.2	

TABLE 2

Gene	Marker	Delta (light/dark)
UNKNOWN (1*)	2142	0.275
UNKNOWN (2)	2120	0.247619048
UNKNOWN (3)	2121	0.21547619
OCA2 (4)	1879	0.19924812
OCA2 (5)	1916	0.188095238
OCA2 (6)	1908	0.183333333
UNKNOWN (7)	2145	0.024767802
UNKNOWN (8)	2109	0.164285714
UNKNOWN (9)	2130	0.154761905
OCA2 (10)	1869	0.114285714
UNKNOWN (11)	2131	0.15
UNKNOWN (12)	2114	0.008333333
UNKNOWN (13)	2128	0.082894737
OCA2 (14)	1905	0.021929825
TYRP (15)	1948	0.078947368
UNKNOWN (16)	2133	0.062656642
UNKNOWN (17)	2110	0.136904762
UNKNOWN (18)	2134	0.061403509
V2-Unknown (19)	1979	0.128571429
UNKNOWN (20)	2137	0.086904762
V2-Unknown (21)	1986	0.128571429
UNKNOWN (22)	2146	0.057894737
UNKNOWN (23)	2147	0.136904762
UNKNOWN (24)	2149	0.055952381
UNKNOWN (25)	2201	0.129761905
UNKNOWN (26)	2151	0.041666667
UNKNOWN (27)	2176	0.128571429
UNKNOWN (28)	2192	0.051587302
UNKNOWN (29)	2166	0.126190476
UNKNOWN (30)	2196	0.042857143
UNKNOWN (31)	2168	0.112781955
OCA2 (32)	1867	0.04047619
OCA2 (33)	1887	0.112573099
TYRP1 (34)	1877	0.107142857
V2-Unknown (35)	1991	0.101190476

*- Sequence Identifier (SEQ ID NO:)

TABLE 3

MARKER*	NAME	MARKERSEQ	ALL1	ALL2
1991	TYRP1E6_354_886938_GA	ATGCTGGTAAGACATTTTCATATGC **C** TTTTGCATGCTCAGCTGGGCRGATT	C	T
1877	TYRP1-rs683_GT	TGAAAAGTATGCCTAGAACTTTAAT **G** CTAACATATGCTTGTATTAGAAAGA	G	T
1887	OCA2-rs2311470_GC	ACTTTTTCATTTAATTTATTAACTT **G** ATGTTTTTAAAGCCTCACAATAATT	G	C
1867	OCA2-rs1375170_GT	AGGGGTCTCTCAGCTAAATCCATAC **G** CTTGCTGAGGCCATGGGGGAACCAA	G	T
2163	AEY-RS920399_TC	AATCTGGATAATTAAAAAATATCTC **C** GAAACTTCACTTAGACCCACACAGA	T	C
2193	revAEY-RS869537_TC	CATTCTCATGGGTTTTGTGAGTGTA **C** GCATGCTTTAGAAGAGCTCTTAGTT	T	C
2168	AEY-RS1036756_TC	AGTTATAGACTAGAAATGAGCCCAG **C** TGAACACTCTATATGAAAAAGACAA	T	C
2192	AEY-RS1464074_TC	NNNNNNNNCTGTCAATACTACTACT **C** TGAATTTGAGAATTTTGAGGAATTT	T	C
2178	(13)AEY-RS1324891_TC	TTCAGTGAATAATCAATTAGGACTA **C** AGTTTCTTGATGTATCTAGAGCATA	T	C
2161	AEYlow-RS725253_TC	AGTGCCACTTATCACCTTGAACATG **C** GCACAGAAAGGGCCTGGCCTATGCT	T	C
2201	AEY-RS2106687_GT	CTAAAGATGAGAAATTGTATTCATT **G** GTTTTCATTGTTAGTTCAGCAGCTG	G	T
2149	AEY-RS951530_GA	AATAGGAGTTAACACAGACCCAGTC **A** TTGAAGCAAGCACCTTGGTAAACCT	G	A
2147	revAEY-RS1386934_GA	AAGTGAAGGTCCAAGAACAAATATG **A** GTGCACCAAACAAACTAACATCACT	G	A
2146	AEY-RS917117_GA	TGAATAACAGAGTCTTTGTCCTTAC **A** GAGTTTGTGGTCTAGTTAGGGAGAN	G	A
1968	MATPex1asn122ser_GA	CTCGTGGGCATGGCTCTGTACCTCA **A** TGGGGCTACTGTTGTAGCAGGTAAG	G	A
2137	AEYlow-RS200393_GA	TATCTTTTGTGCATACACTTTACCC **A** TAATTTCCAGGTGGATTTCTTTGAA	G	A
1979	ASIPrs2424984_ga_GA	GCAGGCACTGGGATCTGCCTGTGAC **A** CTCTTGTGGGGCCTGGTCCCTCGAC	G	A
2134	AEYlow-RS929245_GA	CTATGATATATCGGCATTGTGAAAA **A** ATAGATGGAAGGAGATGACTTGCTG	G	A
2110	AEY-RS764187_GC	TTGCTGCGGCAAGGCTTTCAGCACT **G** AATCAGCTCTCTCCCTCTCCAAGAC	G	C
2133	AEYlow-RS951428_GA	TTGCTGCCAAATATATAACCATACC **A** ACAACTGCTCCACTAATTTTTCAAT	G	A
1948	TYRP_3_2174854_CA	TCTTTCTAATACAAGCATATGTTAG **A** ATTAAAGTTCTAGGCATACTTTTCA	C	A
1905	OCA2-rs1004611_TC	ATTACAATAATGCAAAGTGCATCAA **C** GTCAGAAAGGTGCACCAGAGGCTGG	T	C
2128	AEY-RS1474945_GA	CTGAGTTGGCTTGAGCATGCCTTAT **A** TTCCAGCGAGGGCAAGTAGAATAGA	G	A
2114	AEYlow-RS1821892_GC	CATTTGGAACATAGACATTATAGAT **G** ACAGGCATGTTTTCTTAATCACTAT	G	C
2191	AEY-RS956062_TC	CATTATTTGTTTGTATAGCAGTTTG **C** ACAAATCAAGATCCAAACAAGGGCA	T	C
1869	OCA2-rs1874835_GT	TCAGGCAAACTCCAGAGGAACAGGG **G** AGACGATCGCTTCTGCATTTGCAAA	G	T
2130	AEY-RS1924605_GA	GTCTTTGGGTGCTTTGGATATTTGG **A** TTCCAGTCCTTTATGTGAATTTCTG	G	A
2109	AEY-RS1410591_GC	ATGAGGACAGAGTGTCTCCCTCAGG **G** ATGTGACTTTTCTATGGCAGGCCCC	G	C

TABLE 3 (cont.)

MARKER	NAME	MARKERSEQ	ALL1	ALL2
2145	(20)AEY-RS676290_GA	ATTTATGGAACCCTTTGATTGTATA **A** TGAATCAACATTTTATCCACTAATA	G	A
1908	OCA2-rs895829_TC	CCCCTTACTCAGGCACGTAAGACTC **C** GTGAGCAAGTATAACGCAAGGTGAA	T	C
1916	OCA2-rs1498519_CA	ACTGCCTAGTGTCCTCACTAAGCAA **A** CCCTCAGGACACAGGTTCCAACTCA	C	A
1879	OCA2-rs895828_GC	TGGACAGGTGAGTTGCTGGCCTATC **G** CCTTACTCAGGCACGTAAGACTCCG	G	C
2121	(3)AEY-RS4131568_GA	TGCAACTCAAGAACTCTGATAAACA **A** TATGTTTGGCATTCTGTTCCTGGAA	G	A
2190	AEY-RS1352123_TC	AAAAAGCAGCACTATTATAAAATGA **C** GACATATTCAATGCTGAACAAAGAC	T	C
2142	AEY-RS1934189_GA	TTCCTTTGTATTCTGTAGGCAATTT **A** GATGATGAAATTTAATATTATACCC	G	A

* "MARKER" indicates a SNP identification number.
"NAME" indicates the gene and SNP name.
"MARKSEQ" shows the SNP site (bold), and gene sequences flanking the SNP.
"ALL1" and "ALL2" show the nucleotide occurrence for Allele 1 and Allele 2, respectively, of the gene.

What is claimed is:

1. A method for inferring eye color of a human subject from a nucleic acid sample of the subject, comprising identifying in the nucleic acid sample at least eye color related SNP as set forth in Table 1 or Table 3, whereby the nucleotide occurrence of the SNP is associated with eye color, thereby inferring eye color of the subject.

2. A composition for inferring eye color of a human subject, comprising a specific binding pair member that selectively binds to a polynucleotide comprising a nucleotide occurrence of a SNP as set forth in Table 1 or Table 3, or a polypeptide encoded thereby.

METHODS AND COMPOSITIONS FOR INFERRING EYE COLOR

ABSTRACT OF THE DISCLOSURE

Methods for inferring eye color of an individual from a nucleic acid sample of the individual by detecting the nucleotide occurrence of an eye color related single nucleotide polymorphism (SNP) are provided. Methods for inferring eye color of an individual from a protein sample of the individual by detecting an amino acid residue encoded by the nucleotide occurrence of an eye color related single nucleotide polymorphism (SNP) also are provided. In addition, compositions, including oligonucleotides and antibodies, useful for practicing such methods are provided.

FIGURE 1

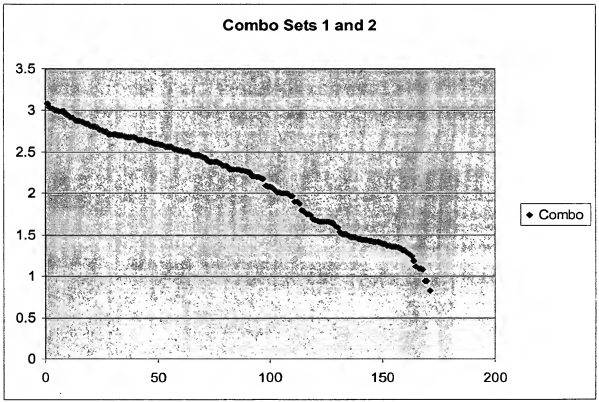
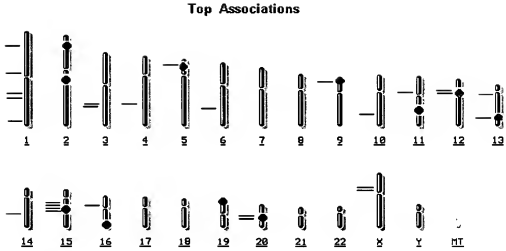


FIGURE 2


Top Associations
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
MT